19TH INTERNATIONAL CONFERENCE ON

ADVANCED INFORMATION NETWORKING AND APPLICATIONS

AINA2005

TAIPEI, TAIWAN
28–30 MARCH 2005

IEEE COMPUTER SOCIETY

IEEE

Edited by
Timothy K. Shih, Tamkang University, Taiwan
Yoshitaka Shibata, Iwate Prefectural University, Japan

Sponsored by
IEEE Computer Society Technical Committee on Distributed Processing (TCDP)
Tamkang University, Taiwan

In cooperation with
Asian Office of Aerospace Research and Development (AOARD)
U.S. Army Asian Research Office

Scalefactor based bit shift FGS audio coding

Fang-Chu Chen and Te-Ming Chiu
Computer & Communications Research Laboratories
Industrial Technology Research Institute
fcchen@itri.org.tw

Abstract

In this paper, a method for applying psychoacoustic effects into the reformation of an MPEG-4 Bit Sliced Arithmetic Coded (BSAC) scalable audio bitstream is presented. The method is to rearrange the compressed information based on the scalefactors calculated by the psychoacuostics model in order to reflect the different subjective significance of the compressed data. At low bit rate it improves the coding efficiencyby a considerable amount.

1. Introduction

While FGS video coding has been touting a lot of research worldwide, FGS audio coding research is limited. This could be due to the fact that the amount of audio data is small compared with that of the video. Samsung's BSAC [1] based MPEG-4 audio coding is one well-designed FGS audio coder. BSAC borrows the bit plane slicing concept from FGS video coding, along with noiseless arithmetic coding, provides a scalable bit stream with granularity as small as 1kbps. However, adding FGS feature costs the coding efficiency of BSAC at low bit rate when not many enhancement layers are received. One good reason is that, while the quantization error of a typical audio coder is controlled by a psychoacoustic model at each specific bit rate, the error introduced by truncating an FGS bit stream is not. If there is a mechanism that the error due to the discarded bits can be governed by the same psychoacoustic model then the coding efficiency can be improved. In this paper we propose such a mechanism termed "scalefactor based bit shift (SFBBS)" which incorporates the influence of the psychoacoustic model in the making of an FGS bit stream.

2. SFBBS

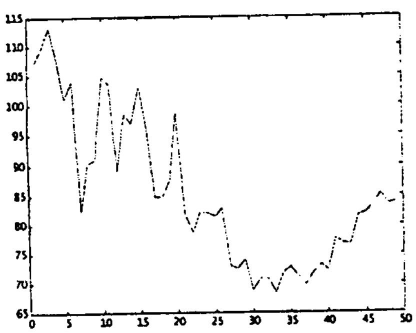
2.1 Scalefactors and the psychoacoustics models

It is well-know that typical audio coding uses psychoacoustic model to keep the compression noise under a masking level so that human ears will not perceive. In MPEG I/II Layer I/II [2][3] audio coders the psychoacoustic model is reflected mostly in the bit rate allocation in each sub-band, scalefactors are used mainly to normalize the dynamic range of each scalefactor band. As the coding technology migrates to MP3 [2] and AAC MPEG-II[4], the scalcfactor finds an important role in phsychoacoutic model's noise shaping process. In MP3[5] or AAC, scalefactors of each sub-bands are used to amplify the signals so that the quantization error can be reduced when the signals are de-amplified at the receiving end. Therefore, if the noise tolerance of a sub-band is small (determined by the psychoacoustic model) the scalefactor will be big so as to keep the quantization noise low. Figure 1. shows the relationship between the scalefactors and the masking curves of two MPEG-4 AAC [6] coded frames. One can note that at those sub-bands where the masking level is smaller the value of their scalefactor is higher. It is this relationship that we relates our scalefactor based bit shift technology to for improving the decoded audio quality at low bit rates of BSAC based coding.

2.2 Scalefactor based Bit shift (SFBBS)

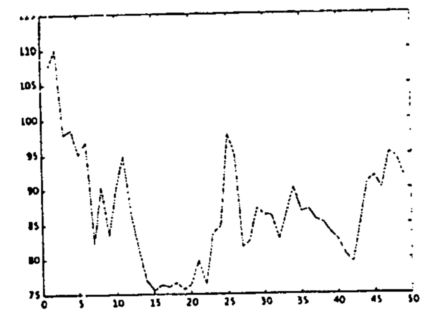
We have to be very careful when saying "keep the error under a masking level so that human ears will not perceive". This is true only for high bit rate audio coding. For low bit rate coding, the error is still perceivable, and the psychoacoustic model in the encoder is only trying to keep the perceivable error as small as possible. For a given bit rate the psychoacoustic model is used in the encoding

band	scalefactor	band	scalefactor
1	0	25	2
2	0	26	0
3	0	27	3
4	0	28	3
5	0	29	2
6	0	30	5
7	2	31	5
8	2	32	5
9	0	33	5
10	0	34	4
11	0	35	4
12	0	36	4
13	1	37	4
14	0	38	4
15	0	39	3
16	0	40	X
17	1	41	X
18	0	42	X
19	0	43	X
20	0	44	X
21	0	45	X
22	2	46	X
23	1	47	X
24	1	48	X
		49	X



(a)

band	scalefactor	band	scalefactor
1	0	25	0
2	0	26	0
3	0	27	1
4	0	28	1
5	0	29	0
6	1	30	0
7	2	31	0
8	0	32	1
9	0	33	0
10	0	34	0
11	1	35	0
12	0	36	0
13	0	37	0
14	3	38	0
15	1	39	0
16	3	40	X
17	4	41	X
18	3	42	1
19	2	43	X
20	3	44	X
21	1	45	X
22	3	46	0
23	0	47	0
24	1	48	X
		49	X



(b)

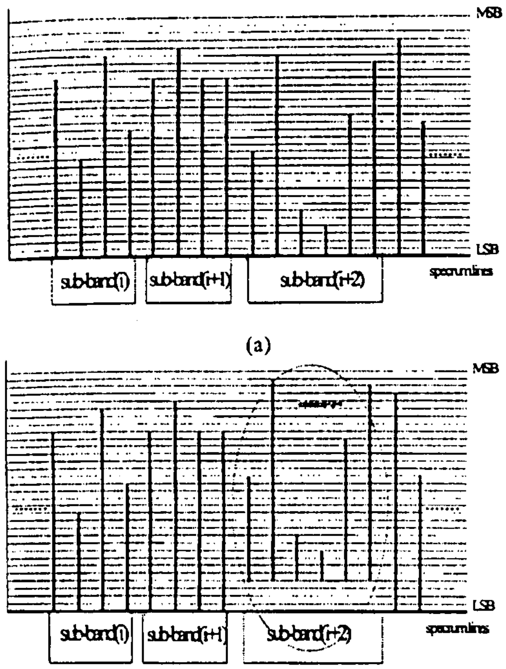
Figure 1. The relationship between the scalefactors and the masking curves (a), (b) for two audio frames

processing to best shape the noise curve. When the bit rate is changed the bit rate allocation algorithm is usually performed again to achieve the best quality at that new bit rate. However, for an FGS coding when the actual received bit rate can not be foreseen by the encoder, running the bit rate allocation algorithm for each possible bit rate is not practical. SFBBS is thus designed to resolve this issue.

What SFBBS does is to up shift the bits of the spectral lines in a scalefactor band according to the corresponding scalefactor before the bit slice process begins. This bit shift concecpt is borrowed from video coding in that when a value is up-shifted its level of importance in the bit slice process is increased. Recall that the scalefactors reflect the psychoacoustical behavior of the signal in the currently processed audio frame, and the bands with less error tolerance are usually associated with bigger scalefactors. By SFBBSing each spectral line we essentially reorder the bits according to their psychoacoustical importance. To be more specifically, a sub-band with small error tolerance indicates that human ears are more sensitive to the frequency range defined by that sub-band. The fact that such sub-bands are with larger scalefactor values allows us to shift the spectral lines in this sub-bands by more bit planes to increase their significance in a greater extent than others. This way we can place those more significant bits closer to the beginning of an FGS bit stream and send them out earlier. In other words, when the FGS bit stream is truncated, those bits which are psychoacoutically less important will be discarded first.

Figure 2. illustrates an example of SFBBS of a sub-band in the case one decides to shift the spectrum in a sub-band by the same number of bit planes as the value of the sub-band's scalefactor. Of course one does not have to shift by the same number as the

corresponding salefactor. As long as the sub-bands with greater scalefactors are shifted no less than those of smaller scalefactors the spirit of SFBBS will be rationed.



(a)

(b)

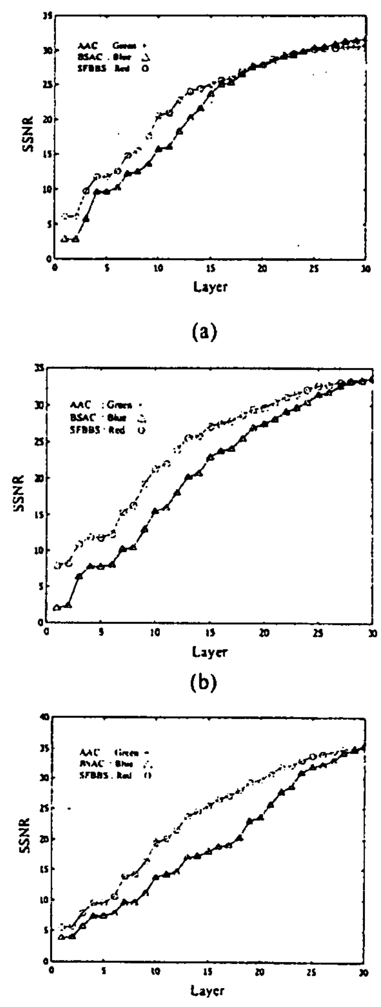
Figure 2. SFBBS of sub-band (i+2) (a) before shift; (b) after shift

2.3. Coding after SFBBS

After SFBBS is performed on the spectral lines BSAC can then follow. However, when performing BSAC on the SFBBSed spectral data one should note that: when a scalefactor band is up shifted the lowest bit planes of that band will carry no meaningful bits and may be excluded from the coding procedure. Let's take Figure 2. as an example, since the band (i+2) shown in Figure 2. is up shifted by four bits, the space in last four bit planes becomes vacant. So when the coding is performed one should skip that space and go on to the next band with meaningful data. This won't cause any confusion in the decoder. Since the decoder has full knowledge of the scalefactors and knows how the spectral lines are up shifted as well as what bit planes are skipped during encoding process, the decoder can do the exact reverse process to restore the original spectral values. By skipping the vacant spaces caused by bit shift, the total bits for coding stay the same with or without SFBBS.

3. Result

Figure 3. shows the performance of SFBBS+BSAC compared with Orginal AAC coding and BSAC only coding. Note that this scalefactor based bit shifting method can improve the audio quality at low bit rate as much as 3 dB and up.



(a)

(b)

(c)

Figure 3. Comparison of coding efficiency among AAC, BSAC only and BSAC+SFBBS

4. Conclusion

In this paper we propose a simple but very powerful pseudo-psychoacoustical noise shaping method used in the making of an BSAC scalable bit stream. As we demonstrate above SFBBS improves the performance of BSAC's coding efficiency. We use the scalefactors to reflect the psychoacoutics model computed in the encoder. Since the scalefactors are sent to the decoder the decoder can perform exact reverse shift to the spectral data. In MPEG-4 SFBBF can be added as a tool before the BSAC block in the encoder and after the BSAC block in the decoder is performed to improved the coding efficiency. It is simple because one only needs minimum efforts to convert BSAC into SFBBS-BSAC. There is no need to change the file format, and no extra overhead is introduced.

5. Reference

[1] ISO/IEC 14496-3, "Information Technology-Coding of Audiovisual Objects: Part 3: Audio Amd. 1".

[2] ISO/IEC 11173-3, "Information Technology-Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s : Part 3: Audio".

[3] ISO/IEC 13818-3, "Information Technology-Generic coding of moving pictures and associated audio information: Part 3: Audio".

[4] ISO/IEC 13818-7, "Information Technology-Generic coding of moving pictures and associated audio information: Part 7: Advanced Audio Coding (AAC)".

[5] F. Baumgarte, C. Feredikis, H. Fuchs, "A Non-Linear Psychoacoustic Model Applied to the ISO MPEG Layer III Coder", Preprint 4087, October 1995.

[6] ISO/IEC 14496-3, "Information Technology-Coding of Audiovisual Objects: Part 3: Audio".